Controlled Query Evaluation under Epistemic Dependencies: Algorithms and Experiments

Abstract. We investigate Controlled Query Evaluation (CQE) over ontologies, where information disclosure is regulated by epistemic dependencies (EDs), a family of logical rules recently proposed for the CQE framework. In particular, we combine EDs with the notion of optimal GA censors, i.e. maximal sets of ground atoms that are entailed by the ontology and can be safely revealed. We focus on answering Boolean unions of conjunctive queries (BUCQs) with respect to the intersection of all optimal GA censors—an approach that has been shown in other contexts to ensure strong security guarantees with favorable computational behavior. First, we characterize the security of this intersection-based approach and identify a class of EDs (namely, full EDs) for which it remains safe. Then, for a subclass of EDs and for DL-Lite$_{\mathcal{R}}$ ontologies, we show that answering BUCQs in the above CQE semantics is in AC^0 in data complexity by presenting a suitable, detailed first-order rewriting algorithm. Finally, we report on experiments conducted in two different evaluation scenarios, showing the practical feasibility of our rewriting function.

Keywords: Description Logics · Ontologies · Confidentiality Preservation · Query Answering · Data Complexity · Epistemic Dependencies.

1 Introduction

The ever-growing volume of structured and semantically rich data has created new challenges for knowledge management and data security. Several modern applications in domains like healthcare and finance rely on ontologies to establish shared vocabularies and formal semantics, facilitating effective data organization and retrieval. While these tools provide sophisticated querying and inference, they also raise critical concerns about information disclosure: sensitive facts may be unintentionally revealed through apparently harmless queries when the underlying ontological axioms are taken into account. *Controlled Query Evaluation* (CQE) [4,5,7,18] is a framework that addresses these concerns by mediating access to data in such a way that only information compliant with a formal *data protection policy*—expressed in logical terms—is accessible through queries.

This work applies CQE to ontologies based on Description Logics (DLs)[1] [2], a family of logics, many of which expressible in first-order (FO) logic, for which the most important reasoning problems are usually decidable. In DLs, knowledge is structured into a TBox, containing intensional axioms, and an ABox, containing extensional facts. The pivotal notion in CQE is the one of *censor*, a set of logical

[1] For an up-to-date overview of CQE in the context of DLs, we refer the reader to [6,13].

formulas that are logically implied by the ontology and comply with the policy. Specifically, we are interested in *GA censors* [15], which consist of ground atoms, hence structurally resembling ABoxes or relational databases.

Usually, CQE policies are defined using *denials*, i.e. expressions of the form $(\exists \mathbf{x}\, \gamma(\mathbf{x})) \rightarrow \bot$, where $\exists \mathbf{x}\, \gamma(\mathbf{x})$ is a *Boolean conjunctive query* (BCQ). Denials are used to specify information that must remain undisclosed: the system must prevent users from inferring that the formula $\exists \mathbf{x}\, \gamma(\mathbf{x})$ holds in the ontology. In the recent work [12], the authors proposed an extension for such a language of rules, called *epistemic dependencies* (EDs) [16], which are logical implications between two (possibly open) conjunctive queries, each within the scope of a modal operator K, though adopting a notion of censor that differs from the one considered in the current paper.

Example 1. The policy of a company stipulates that all salaries of employees must be kept confidential, except those of managers. In addition, the existence of consensual personal relationships between managers and their employees must remain undisclosed.

In logical terms, such a policy can be defined as the following set of EDs:

$$\begin{aligned}\mathcal{P} = \{\, & \forall x, y\, (K\mathsf{salary}(x, y) \rightarrow K\mathsf{manager}(x)),\\ & K\exists x, y\, (\mathsf{managerOf}(x, y) \wedge \mathsf{consRel}(x, y)) \rightarrow K\bot \,\}\end{aligned}$$

where $\mathsf{manager}$ is a unary predicate indicating that an individual is a manager, and salary, $\mathsf{consRel}$ and $\mathsf{managerOf}$ are binary predicates modelling, respectively, the salary level of a person, the consensual relationship between two individuals and the relationship where one individual manages another. In the second ED, the usage of the existential quantifier indicates that, for every manager (resp., employee), the very existence of a consensual relationship with any employee (resp., manager) of hers must not be revealed—not merely the identities of the individual involved.

Moreover, suppose that the company ontology consists of:

- A TBox $\mathcal{T} = \{\exists\mathsf{managerOf} \sqsubseteq \mathsf{manager}, \mathsf{manager} \sqsubseteq \exists\mathsf{respDept}\}$, meaning that everyone who manages another individual is a manager, and managers are such only if they are responsible for some department.
- An ABox $\mathcal{A} = \{\mathsf{managerOf}(\mathsf{lucy}, \mathsf{tom}), \mathsf{consRel}(\mathsf{lucy}, \mathsf{tom}), \mathsf{salary}(\mathsf{lucy}, \mathsf{150k}), \mathsf{salary}(\mathsf{tom}, \mathsf{75k})\}$, meaning that Lucy is Tom's manager, they have a consensual relationship, and their salary is \$150,000 and \$75,000, respectively.

A censor consisting only of ground atoms must remove at least one of the facts $\mathsf{managerOf}(\mathsf{lucy}, \mathsf{tom}), \mathsf{consRel}(\mathsf{lucy}, \mathsf{tom})$ from $\mathcal{A}$ and, at the same time, must remove the fact $\mathsf{salary}(\mathsf{tom}, \mathsf{75k})$ (because Tom is not a manager). By contrast, any optimal GA censor can safely include the facts $\mathsf{manager}(\mathsf{lucy})$ and $\mathsf{salary}(\mathsf{lucy}, \mathsf{150k})$, because Lucy's managerial status follows from the ontology and knowing that she is a manager (and her salary) does not violate the policy. ■

Our main objective is to evaluate queries under a formal entailment semantics that maximizes data disclosure while remaining compliant with the policy. Thus, we call *optimal GA censors* the GA censors that are maximal w.r.t. set inclusion, and focus on the problem of checking whether a *Boolean union of conjunctive queries* (BUCQ) is entailed by the TBox and the intersection of all the optimal GA censors. This task, known as *IGA-entailment*, has been shown to be *FO-rewritable* when the TBox is expressed in DL-Lite$_{\mathcal{R}}$ and the policy consists of denials [14]. That is, in the above setting, IGA-entailment of a BUCQ q can be decided by rewriting q into a new FO query q_r that only depends on the TBox and the policy and, in a second moment, evaluating q_r over the ABox. This property guarantees a nice computational behaviour at a theoretical level, as the task enjoying it has the same complexity as evaluating an SQL query over a database. However, it still needs to be empirically validated through a practical implementation. In the case based on denials, a working prototype was provided in [11], within the ontology-based data access framework.

We aim to extend this scenario to accommodate policies defined using EDs while preserving the *FO-rewritability* property. First, we prove that the class of *full EDs* and *linear EDs* enjoy a desirable property related to security. We exclude, however, the possibility that IGA-entailment remains FO-rewritable for such classes of dependencies, by proving coNP- and NL-hardness results for the related decision problem, respectively. We thus identify a condition for full EDs for which we are able to prove the FO-rewritability. For two classes of EDs that respect such a condition, namely the linear full and the acyclic full EDs, we finally conducted experiments to test the practical feasibility of our rewriting algorithm. Specifically, we implemented a tool that rewrites a SPARQL BUCQ into a new query q_r solely based on the given TBox and policy, and then evaluates q_r over an SQL database containing the ABox. Since our theoretical results are related to the logic DL-Lite$_{\mathcal{R}}$, we adopted the OWL 2 QL ontology of the OWL2Bench benchmark [24] as our testbed. Two distinct evaluation scenarios demonstrate that our method is not only theoretically sound but also practically feasible, with most rewritten queries running within acceptable time bounds.

The paper is structured as follows. Section 2 provides the necessary theoretical background. Section 3 describes the framework and the problem studied. Section 4 introduces a key property that helps identify interesting subclasses of EDs. Section 5 presents lower bounds that exclude FO-rewritability in the general case. Section 6 defines the subclass of EDs we target and presents a detailed FO-rewriting algorithm. Section 7 reports our experimental findings. Finally, Section 8 concludes the paper.

2 Preliminaries

In this paper, we refer to standard notions of function-free first-order (FO) logic and Description Logics (DL). We use countably infinite sets of symbols Σ_C, Σ_R, Σ_I and Σ_V, containing respectively unary predicates (called *concepts*), binary predicates (called *roles*), constant symbols (also called *individuals*) and variables. An *atom* is a formula of the form $P(\mathbf{t})$, where P is a (either unary or binary) predicate and $\mathbf{t}$ is a sequence of terms, i.e. variables or constants.

Given a set of FO formulas Φ, we denote by $vars(\Phi)$ and $pred(\Phi)$, the sets of variables and predicates occurring in Φ, respectively. Given any FO formula ϕ, we use the notation $\phi(\mathbf{x})$ when we want to emphasize its free variables $\mathbf{x}$, and we overload $vars(\cdot)$ to work with FO formulas, with the same meaning. If $\phi(\mathbf{x})$ is closed (that is, $\mathbf{x}$ is empty), then it is called a *sentence*; Furthermore, if $vars(\phi) = \emptyset$, then ϕ is said to be *ground*. In particular, ground atoms are also called *facts*. If $\mathcal{F}$ is a set of facts, we say that an FO sentence ϕ evaluates to true in $\mathcal{F}$ to actually mean that it is true in the Herbrand model of $\mathcal{F}$.

In this work, we also refer to specific classes of domain-independent FO formulas, such as the class of *conjunctive queries* (CQs), i.e. formulas of the form $\exists \mathbf{x}\,(\gamma)$, where $\mathbf{x} \subseteq vars(\gamma)$ and γ is a conjunction of atoms. A disjunction of CQs sharing the same free variables is called *union of conjunctive queries* (UCQ), which sometimes we also treat as a set of CQs. As customary, closed CQs and UCQs are said to be *Boolean* and referred to as BCQs and BUCQs, respectively. Given any CQ q, we indicate by $QA(q)$ the set of atoms of q.

We call *substitution* any function $\sigma : \Sigma_V \to \Sigma_V \cup \Sigma_I$. Given a CQ q and a substitution σ of (a subset of) its variables, we write $\sigma(q)$ to denote the result of applying σ to q. A substitution of variables is said to be *ground* if its image is contained in Σ_I. Moreover, given a set of facts $\mathcal{F}$, an *instantiation* for q in $\mathcal{F}$ is a ground substitution σ of the variables of q such that $QA(\sigma(q)) \subseteq \mathcal{F}$. If such an instantiation σ exists, the set $QA(\sigma(q))$ is called *image* of q in $\mathcal{F}$. Given a UCQ q, an *image of q in $\mathcal{F}$* is any image of q' in $\mathcal{F}$, for any $q' \in q$.

We resort to DL ontologies as a formal way of representing structured knowledge about a given domain. Ontologies are usually partitioned into two sets used, respectively, for representing intensional and extensional knowledge. More formally, given a DL $\mathcal{L}_\mathcal{T}$, an $\mathcal{L}_\mathcal{T}$ *ontology* is a finite set $\mathcal{T} \cup \mathcal{A}$, where $\mathcal{T}$ (called $\mathcal{L}_\mathcal{T}$ TBox) is a set of axioms expressible in $\mathcal{L}_\mathcal{T}$, and $\mathcal{A}$ (called ABox) is a set of ground atoms. In particular, our complexity results hold for ontologies expressed in DL-Lite$_\mathcal{R}$ [10], which is the logic underpinning OWL 2 QL, one of the three OWL 2 profiles [17,21] that is specifically designed for efficient query answering. The axioms of a DL-Lite$_\mathcal{R}$ TBox $\mathcal{T}$ take the following form:

$$B \sqsubseteq B', \quad R \sqsubseteq R', \quad B \sqsubseteq \neg B', \quad R \sqsubseteq \neg R'$$

where B and B' (resp., R and R') are of the form A, $\exists S$ or $\exists S^-$ (resp., of the form S or S^-), with $A \in \Sigma_C$, $S \in \Sigma_R$ and S^- the inverse of S. The unqualified existential restriction $\exists S$ (resp., $\exists S^-$) represents the set of individuals occurring as the first (resp., second) argument of S.

Moreover, we denote with $\mathsf{cl}_\mathcal{T}(\mathcal{A})$ the *closure* of $\mathcal{A}$ w.r.t. $\mathcal{T}$, i.e. the set of all the ground atoms that are logical consequences of $\mathcal{T} \cup \mathcal{A}$. We also refer to the rewriting function $\mathsf{AtomRewr}$, for which the following property has been demonstrated in [15, Lemma 6].

Proposition 1 ([15]). *Let $\mathcal{T} \cup \mathcal{A}$ be a consistent DL-Lite$_\mathcal{R}$ ontology, and let ϕ be an FO sentence. Then, ϕ evaluates to true in $\mathsf{cl}_\mathcal{T}(\mathcal{A})$ if and only if $\mathsf{AtomRewr}(\phi, \mathcal{T})$ evaluates to true in $\mathcal{A}$.*

Finally, Section 5 requires the reader to be familiar with basic notions of computational complexity theory [22]. Indeed, decision problems solvable through FO-rewriting are known to be in AC^0 in data complexity [1], and we refer to the complexity classes NL and coNP (both of which are known to be strict supersets of AC^0) for showing scenarios in which the studied problem is not FO-rewritable.

3 Framework

In the spirit of [12], we adopt as *protection policy* (or simply *policy)* a finite set of epistemic dependencies, which are a special case of EQL-Lite(CQ) [9] sentences, and are defined as follows.

Definition 1. *An* epistemic dependency (ED) *is a sentence τ of the following-form:*

$$\forall \mathbf{x}_1, \mathbf{x}_2 \, (Kq_b(\mathbf{x}_1, \mathbf{x}_2) \rightarrow Kq_h(\mathbf{x}_2)) \tag{1}$$

where $q_b(\mathbf{x}_1, \mathbf{x}_2)$ is a CQ with free variables $\mathbf{x}_1 \cup \mathbf{x}_2$, $q_h(\mathbf{x}_2)$ is a CQ with free variables $\mathbf{x}_2$, and K is a modal operator.

We say that a FO theory Φ *satisfies* an ED τ (in symbols $\Phi \models_{\mathsf{EQL}} \tau$) if, for every ground substitution σ for the free variables of $\mathsf{body}(\tau)$, if $\Phi \models \sigma(\mathsf{body}(\tau))$ then $\Phi \models \sigma(\mathsf{head}(\tau))$. If all the EDs of a policy $\mathcal{P}$ are satisfied by Φ, then we say that Φ *satisfies* $\mathcal{P}$ (in symbols $\Phi \models_{\mathsf{EQL}} \mathcal{P}$). Intuitively, EDs express the disclosure rules that should govern the publication of data. For every ground substitution σ of the universal variables of an ED τ, if the ontology entails $\sigma(\mathsf{body}(\tau))$, then this information may only be disclosed if $\sigma(\mathsf{head}(\tau))$ can also be disclosed. In the special case in which $\mathsf{head}(\tau) = \bot$, the ED acts as a denial constraint, i.e. its body must not be revealed under any instantiation of its free variables.

The input to our framework is a triple $\langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$, that we call *CQE instance*, where $\mathcal{T}$ is a DL-Lite$_{\mathcal{R}}$ TBox, $\mathcal{P}$ is a policy and $\mathcal{A}$ is an ABox. Given a CQE instance $\langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$, it may be the case that $\mathcal{T} \cup \mathcal{A} \not\models_{\mathsf{EQL}} \mathcal{P}$, i.e. the policy is not satisfied by the ontology $\mathcal{T} \cup \mathcal{A}$. For hiding the part of information that should be protected, we rely on the following notion of GA censor [15].

Definition 2 (GA censor). *Given a CQE instance $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$, a* ground atom censor *(in short,* GA censor*) for $\mathcal{E}$ is any subset $\mathcal{C}$ of $\mathsf{cl}_{\mathcal{T}}(\mathcal{A})$ such that $\mathcal{T} \cup \mathcal{C} \models_{\mathsf{EQL}} \mathcal{P}$. A GA censor $\mathcal{C}$ of $\mathcal{E}$ is* optimal *if no other GA censor $\mathcal{C}'$ of $\mathcal{E}$ exists such that $\mathcal{C}' \supseteq \mathcal{C}$.*

Informally, an optimal GA censor is a maximal (w.r.t. set inclusion) portion of $\mathsf{cl}_{\mathcal{T}}(\mathcal{A})$ that can be safely disclosed without violating the given policy. Among the possibly many GA censors for a CQE instance, we are particularly interested in the information that is common to all optimal GA censors, which represents the maximal amount of facts that can be safely revealed in every scenario where policy satisfaction is preserved. Based on this idea, we now introduce the central notion of IGA-entailment, which formalizes query answering under the intersection of all optimal GA censors.

Definition 3 (IGA-entailment). *Given a CQE instance $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$ and a BUCQ q, we say that $\mathcal{E}$* IGA-entails *q (in symbols $\mathcal{E} \models_{\mathsf{IGA}} q$) if $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \models q$, where $\mathcal{C}_{\mathsf{IGA}}$ is the intersection of all the optimal GA censors of $\mathcal{E}$.*

Example 2. Consider the CQE instance $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$, where $\mathcal{T}$, $\mathcal{P}$ and $\mathcal{A}$ are as in Example 1. In this case, we have only two optimal censors, i.e.:

$$\mathcal{C}_1 = \{\mathsf{managerOf}(\mathsf{lucy}, \mathsf{tom}), \mathsf{manager}(\mathsf{lucy}), \mathsf{salary}(\mathsf{lucy}, \mathsf{150k})\},$$
$$\mathcal{C}_2 = \{\mathsf{consRel}(\mathsf{lucy}, \mathsf{tom}), \mathsf{manager}(\mathsf{lucy}), \mathsf{salary}(\mathsf{lucy}, \mathsf{150k})\}.$$

Now, we have that a BCQ like $q_1 = \exists x, y\, \mathsf{consRel}(x, y)$ is not IGA-entailed by $\mathcal{E}$, because the extension of consRel is empty in $\mathcal{C}_{\mathsf{IGA}} = \mathcal{C}_1 \cap \mathcal{C}_2$ and no relevant conclusions can be drawn from the intensional axioms. Conversely, the BCQ $q_2 = \exists x, y\, (\mathsf{respDept}(x, y) \wedge \mathsf{salary}(x, \mathsf{150k}))$ is IGA-entailed by $\mathcal{E}$, because the facts manager(lucy) and salary(lucy, 150k) belong to the intersection $\mathcal{C}_{\mathsf{IGA}}$. ■

It is worth noting that an alternative, more expressive kind of entailment has been proposed in the literature, i.e., the skeptical entailment over all the optimal GA censors. However, such an approach has already been shown to be computationally intractable (specifically, coNP-hard in data complexity) even for BCQs and for policies restricted to sets of denial constraints [15, Thm. 6].

Although the above definitions apply to any CQE instance, our complexity analysis focuses on specific classes on EDs, which are defined below. First, borrowing the terminology from the literature on both databases and existential rules (see e.g. [3,8]), an ED τ is called *full* if no existential variable occurs in its head, and is called *linear* if $|QA(\mathsf{body}(\tau))| = 1$. A policy is *full* (resp., *linear*) if all its EDs are full (resp., linear). Then, following the paper [12] that introduced EDs for CQE, we define the notion of acyclicity as follows. Given a policy $\mathcal{P}$ and a TBox T, consider the graph G whose nodes are the predicates of $\mathcal{T} \cup \mathcal{P}$, and whose edges are of two kinds: T-edges, which connect two nodes A and B of G if they occur, respectively, on the left- and right-hand side of a concept inclusion of $\mathcal{T}$; analogously P-edges connect two nodes A and B of G if they occur, respectively, in the body and head of an ED of $\mathcal{P}$. Then, $\mathcal{P}$ is said to be *acyclic for $\mathcal{T}$* if G contains no cycle involving a P-edge.

We finally define IGA-Ent as the following decision problem.

Problem:	IGA-Ent
Input:	A DL-Lite$_{\mathcal{R}}$ CQE instance $\mathcal{E} = \langle \mathcal{T}, \mathcal{A}, \mathcal{P} \rangle$, a BUCQ q
Question:	Does $\mathcal{E} \models_{\mathsf{IGA}} q$?

4 Security of the Intersection of GA Censors

Before delving into the complexity analysis of the IGA-Ent problem, it is natural to ask whether the intersection of all optimal GA censors constitutes in turn a valid censor, i.e. whether it satisfies the given policy.

Actually, the above property does not hold, in general, even in the case when the TBox is empty, as shown by the following example.

Example 3. Consider the CQE instance $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$, where $\mathcal{T}$ is empty and

$$\mathcal{P} = \{ \tau_1 : K(B(1) \land B(2)) \rightarrow K\bot, \quad \tau_2 : \forall x\, (KC(x) \rightarrow K\exists y\, B(y)) \}, \qquad \mathcal{A} = \{C(0), B(1), B(2)\}.$$

It is immediate to see that $\mathcal{C}_1 = \{C(0), B(1)\}$ and $\mathcal{C}_2 = \{C(0), B(2)\}$ are the only two optimal GA censors for $\mathcal{E}$, and so their intersection $\mathcal{C}_{\mathsf{IGA}}$ is $\{C(0)\}$. Now, we observe that $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \not\models_{\mathsf{EQL}} \tau_2$, which implies that $\mathcal{C}_{\mathsf{IGA}}$ is not a censor. ■

Notice that the policy employed in Example 3 is acyclic for the coupled (empty) TBox. However, it is possible to demonstrate that it is safe to refer to the intersection of the optimal GA censors when considering the classes of linear EDs and full EDs, which we focus on in the next section.

Theorem 1. *For every CQE instance $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$ such that $\mathcal{P}$ is either full or linear, the set $\mathcal{C}_{\mathsf{IGA}} = \bigcap_{\mathcal{C} \in \mathsf{optCens}(\mathcal{E})} \mathcal{C}$ is a GA censor of $\mathcal{E}$.*

Proof. To be a GA censor of $\mathcal{E}$, $\mathcal{C}_{\mathsf{IGA}}$ must be a subset of $\mathsf{cl}_{\mathcal{T}}(\mathcal{A})$ and be such that $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \models_{\mathsf{EQL}} \mathcal{P}$. The first condition always follows by construction of $\mathcal{C}_{\mathsf{IGA}}$. Then, for both the cases have just to prove the second one.

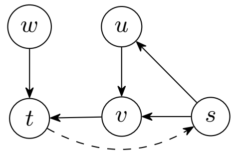
We first focus on full policies. To prove the thesis, let us consider any $\tau \in \mathcal{P}$ and any ground substitution σ of the universal variables of τ such that $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \models \sigma(\mathsf{body}(\tau))$, and let us show that $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \models \sigma(\mathsf{head}(\tau))$. If $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \models \sigma(\mathsf{body}(\tau))$, then by monotonicity, it follows that $\mathcal{T} \cup \mathcal{C} \models \sigma(\mathsf{body}(\tau))$ for every $\mathcal{C} \in \mathsf{optCens}(\mathcal{E})$. However, for every such $\mathcal{C}$, since τ is full, then $\sigma(\mathsf{head}(\tau))$ is a ground conjunction of atoms, and since $\mathcal{C}$ is optimal, then all such ground atoms belong to it. Consequently, all the ground atoms of $\sigma(\mathsf{head}(\tau))$ belong to $\mathcal{C}_{\mathsf{IGA}}$, which immediately implies that $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \models \sigma(\mathsf{head}(\tau))$.

We now turn our attention to linear policies. Notice that, if every ED is linear, then only one optimal GA censor $\mathcal{C}$ for $\mathcal{E}$ exists. Such a GA censor can indeed be deterministically computed by iteratively (until a fixpoint is reached) checking whether there exists an atom $\alpha \in \mathcal{C}$, an ED $\tau \in \mathcal{P}$ and an assignment σ of the free variables of $\mathsf{body}(\tau)$ such that $\mathcal{T} \cup \{\alpha\} \models \sigma(\mathsf{body}(\tau))$ but $\mathcal{T} \cup \{\alpha\} \not\models \sigma(\mathsf{head}(\tau))$ and, in such a case, remove α from $\mathcal{C}$. Thus, the intersection coincides with such a censor, i.e. $\mathcal{C}_{\mathsf{IGA}} = \mathcal{C}$, which implies $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \models_{\mathsf{EQL}} \mathcal{P}$ by Definition 2. □

5 Negative Results

As anticipated in the Introduction, we aim to find a class of dependencies for which IGA-entailment is FO-rewritable. Starting from the results obtained in the previous section, a natural first choice would be the classes of full and linear EDs. The following property, however, excludes the possibility of FO-rewritability of IGA-entailment for the class of linear epistemic dependencies.

Theorem 2. *There exist an empty TBox, a policy consisting of linear EDs, and a query consisting of only one atom for which* IGA-ENT *is NL-hard in data complexity.*



$$\mathcal{A} = \{\mathit{reachableVia}(u,s),\ \mathit{reachableVia}(v,s),\ \mathit{reachableVia}(v,u),\ \mathit{reachableVia}(t,v),\ \mathit{reachableVia}(s,s),\ \mathit{reachableVia}(s,t),\ \mathit{reachableVia}(t,w)\,\}$$

$$\mathcal{C}_{\mathsf{IGA}} = \{\mathit{reachableVia}(u,s),\ \mathit{reachableVia}(v,s),\ \mathit{reachableVia}(v,u),\ \mathit{reachableVia}(t,v),\ \mathit{reachableVia}(s,s),\ \mathit{reachableVia}(s,t)\,\}$$

Fig. 1. A possible instance of the graph $\mathcal{G}$ and the corresponding sets $\mathcal{A}$ and $\mathcal{C}_{\mathsf{IGA}}$, built as in the proof of Theorem 2. The dashed edge only belongs to $\mathcal{G}'$.

Proof. We prove the thesis by showing a logspace reduction from the restriction of st-connectivity to directed acyclic graphs (DAGs)[2], which is known to be NL-hard. Given a DAG $\mathcal{G} = \langle V, E\rangle$ and two nodes $s, t \in V$, such a problem consists of checking whether there exists a path from s to t in $\mathcal{G}$.

First, let us fix the policy as follows:

$$\mathcal{P} = \{\forall u\,(K\exists v\ \mathit{reachableVia}(v,u) \rightarrow K\exists w\ \mathit{reachableVia}(u,w))\}.$$

Now, let $\mathcal{G}' = \langle V, E'\rangle$ be such that $E' = E \cup \{(t,s)\}$, i.e. the graph identical to $\mathcal{G}$ except that it contains a further edge from t to s. Let also $\mathcal{T} = \emptyset$ and

$$\mathcal{A} = \{\mathit{reachableVia}(v,u) \mid (u,v) \in E'\} \cup \{\mathit{reachableVia}(s,s)\}.$$

Note that $\mathcal{A}$ (we ignore the TBox, as it is empty) does not necessarily satisfy the policy, but the set $\{\mathit{reachableVia}(s,s)\}$ does. Intuitively, the policy $\mathcal{P}$ states that, for every pair of nodes $v, u \in V$, v is reachable from s via u only if u is reachable from s via another node $w \in V$. Also, the ABox states that every node is reachable from s via all its predecessors in $\mathcal{G}'$ (which is not necessarily true), plus the fact that s is reachable via itself.

Notice also that, since the unique ED in $\mathcal{P}$ contains only one atom in its body, then there exists exactly one optimal GA censor for $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A}\rangle$, which then coincides with $\mathcal{C}_{\mathsf{IGA}}$.

We can now prove that the following three statements are equivalent:

1. $\mathcal{G}$ contains a path from s to t;
2. there exists a directed cycle in $\mathcal{G}'$;
3. the BCQ $q = \exists u\ \mathit{reachableVia}(t,u)$ is IGA-entailed by $\mathcal{E}$.

($1 \Leftrightarrow 2$) Since $\mathcal{G}$ is a DAG, it is immediate to verify that $\mathcal{G}$ contains a path from s to t iff $\mathcal{G}'$ is cyclic.

($2 \Rightarrow 3$) If a directed cycle exists in $\mathcal{G}'$, then it involves s and t (otherwise $\mathcal{G}$ would not be a DAG). In this case, $\mathcal{A}$ contains the following set of facts:

$$\{\,\mathit{reachableVia}(s,s), \mathit{reachableVia}(u_1,s), \mathit{reachableVia}(u_2,u_1), \mathit{reachableVia}(u_3,u_2), \ldots, \mathit{reachableVia}(t,u_m), \mathit{reachableVia}(s,t)\,\}$$

[2] Such a problem is sometimes referred to as DAG-STCON or STCONDAG.

which is contained in the unique optimal GA censor for $\mathcal{E}$. Thus, we have that $\mathcal{E} \models_{\mathsf{IGA}} q$.

(3 $\Rightarrow$ 1) Suppose that there is no path from s to t in $\mathcal{G}$, i.e. every directed path ending in t does not include s. Also, since $\mathcal{G}$ is a DAG, none of these paths contains a cycle. Given this and the way $\mathcal{A}$ is constructed, it follows immediately that no fact of the form $\mathit{reachableVia}(t, \_)$ can be part of the unique optimal GA censor for $\mathcal{E}$ (which, as stated above, coincides with $\mathcal{C}_{\mathsf{IGA}}$). Indeed, if any fact $\mathit{reachableVia}(t, u)$ were included in $\mathcal{C}_{\mathsf{IGA}}$, the policy would require the inclusion of another fact $\mathit{reachableVia}(u, w)$, and so on, until one eventually reaches a fact (corresponding to the final edge of a path ending at t) beyond which no further fact can be added. Therefore, we have that $\mathcal{E} \not\models_{\mathsf{IGA}} q$. □

We now show that the scenario is even worse for full dependencies, as the decision problem under consideration is even intractable in data complexity.

Theorem 3. *There exist an empty TBox, a policy consisting of full EDs, and a query consisting of a ground atom for which IGA-ENT is coNP-hard in data complexity.*

Proof. We prove the thesis by reduction from 3-CNF. We define the TBox $\mathcal{T} = \emptyset$ and the following full policy:

$$\begin{array}{r}
\forall i, j, p, v\ (K\,(S(i) \wedge N(j,i) \wedge V_1(i,p) \wedge P_1(i,v) \wedge T(p,v)) \rightarrow K\,S(j)) \\
\forall i, j, p, v\ (K\,(S(i) \wedge N(j,i) \wedge V_2(i,p) \wedge P_2(i,v) \wedge T(p,v)) \rightarrow K\,S(j)) \\
\forall i, j, p, v\ (K\,(S(i) \wedge N(j,i) \wedge V_3(i,p) \wedge P_3(i,v) \wedge T(p,v)) \rightarrow K\,S(j)) \\
\forall x\ (K\,(T(x,0) \wedge T(x,1)) \rightarrow K\,U(0))
\end{array}$$

Now, given a 3-CNF formula φ with m clauses, we build the ABox $\mathcal{A}$ as the set of the following:

- $N(i-1, i)$ for every $1 \leq i \leq m$, meaning that c_i is the next clause of c_j (note that $N(0,1)$ is included);
- $V_j(i,p)$ if p is the j-th propositional variable in the i-th clause;
- $P_j(i,0)$ (resp., $P_j(i,1)$) if the polarity of the j-th literal in the i-th clause is negative (resp., positive);
- $T(p,0)$ and $T(p,1)$ for every propositional variable p occurring in φ;
- $S(1), S(2), \ldots, S(m)$.

We prove that φ is unsatisfiable iff $S(m)$ is IGA-entailed by $\mathcal{E}$.

($\Leftarrow$) Suppose φ is satisfiable. Let P be the set of propositional variables occurring in φ, let $I \subseteq P$ be an interpretation satisfying φ, and let $\mathcal{A}'$ be the following subset of $\mathcal{A}$:

$$\begin{aligned}
\mathcal{A}' = \mathcal{A} \setminus (&\{T(p,0) \mid p \in I\} \cup \\
&\{T(p,1) \mid p \in P \setminus I\} \cup \{S(1), \ldots, S(m)\})
\end{aligned}$$

It is immediate to verify that $\mathcal{T} \cup \mathcal{A}' \models_{\mathsf{EQL}} \mathcal{P}$, as it does not contain any S fact and there does not exist any constant c such that $\{T(c,0), T(c,1)\} \subseteq \mathcal{A}'$.

Algorithm 1: PolicyExp

input : A DL-Lite$_{\mathcal{R}}$ TBox $\mathcal{T}$, a policy $\mathcal{P}$ expandable w.r.t. $\mathcal{T}$

$\mathcal{P}' \leftarrow \emptyset$;

foreach $\tau \in \mathcal{P}$ **do**

  foreach $q(\mathbf{x}) \in \mathsf{UCQRew}(\mathsf{body}(\tau), \Sigma(\mathcal{P}, \mathcal{T}))$ **do**

    $\mathcal{P}' \leftarrow \mathcal{P}' \cup \{\forall \mathbf{x}(Kq(\mathbf{x}) \rightarrow K\mathsf{head}(\tau))\}$;

return $\mathcal{P}'$

This implies that $\mathcal{A}'$ is part of at least one optimal GA censor of $\mathcal{E}$ (as it is a GA censor of $\mathcal{E}$ itself). Moreover, one can see that $\mathcal{A}' \cup \{S(m)\}$ is not part of any optimal GA censors of $\mathcal{E}$. This is proved by the fact that the addition of $S(m)$ creates a sequence of instantiations of the bodies of the first three EDs of $\mathcal{P}$ that requires (to preserve the satisfaction of the policy) to add to $\mathcal{A}' \cup \{S(m)\}$ first the fact $S(m-1)$, then $S(m-2)$, and so on until $S(1)$: this would in turn imply to also add $S(0)$, which however, does not belong to $\mathsf{cl}_{\mathcal{T}}(\mathcal{A})$.

Consequently, there exists an optimal GA censor $\mathcal{C}$ of $\mathcal{E}$ that does not contain $S(m)$. Thus, $S(m)$ is not part of the intersection $\mathcal{C}_{\mathsf{IGA}}$ of the optimal GA censors, and since the TBox $\mathcal{T}$ is empty, we directly have that $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \not\models S(m)$, i.e. $\mathcal{E} \not\models_{\mathsf{IGA}} S(m)$.

($\Rightarrow$) Given a guess of the atoms of the T predicate satisfying the fourth dependency and corresponding to an interpretation of the propositional variables that does not satisfy φ, it is straightforward to verify that the sequence of instantiations of the bodies of the first three EDs of $\mathcal{P}$ mentioned above (which has previously lead to the need of adding $S(0)$ to the set) results to be interrupted due to the absence of some fact for T. This missing fact reflects a variable whose truth value conflicts with the polarity of the corresponding literal. Then, there exists a positive integer $k \leq m$ such that the atoms $S(k), S(k+1), \ldots, S(m)$ can be added to all the optimal GA censors corresponding to such a guess of the T atoms. Since this holds for every interpretation of φ (none of which satisfies the propositional formula), then the unsatisfiability of φ implies that $\mathcal{E} \models_{\mathsf{IGA}} S(m)$. □

6 A FO-Rewritable Class of EDs

In this section, we identify a condition for full EDs such that it is possible to solve IGA-entailment through a FO-rewriting technique in the case of DL-Lite$_{\mathcal{R}}$ TBoxes. We start by defining when a given set $\mathcal{F}$ of facts is disclosable. Intuitively, this happens when we can expose it via at least one optimal GA censor.

Definition 4 (Disclosability). *Given a CQE instance $\mathcal{E}$, we say that a set of facts $\mathcal{F}$ is* disclosable in $\mathcal{E}$ *if there exists a subset $\mathcal{F}'$ of $\mathsf{cl}_{\mathcal{T}}(\mathcal{A})$ such that $\mathcal{F} \subseteq \mathcal{F}'$ and $\mathcal{T} \cup \mathcal{F}' \models_{\mathsf{EQL}} \mathcal{P}$.*

For technical purposes, we make use of the notion of *tuple-generating dependency* (TGD) [1], i.e. FO expressions of the form $\forall \mathbf{x}_1, \mathbf{x}_2\, (q_b(\mathbf{x}_1, \mathbf{x}_2) \rightarrow q_h(\mathbf{x}_2))$, where both $q_b(\mathbf{x}_1, \mathbf{x}_2)$ and $q_h(\mathbf{x}_2)$ are CQs. We say that a set Σ of TGDs is *UCQ-rewritable* if, given any CQ $q(\mathbf{x})$, there exists a UCQ $\mathsf{UCQRew}(q, \Sigma)$ such that, for every set $\mathcal{F}$ of facts and for every ground substitution σ of the free variables of q, $\Sigma \cup \mathcal{F} \models \sigma(q)$ iff $\mathcal{F} \models \sigma(q_r)$ for some $q_r(\mathbf{x}) \in \mathsf{UCQRew}(q, \Sigma)$). In this regard, we refer to [3], which establishes a sufficient condition for UCQ rewritability (generalizing acyclicity and linearity), and to [20], which provides a suitable rewriting algorithm.[3]

Now, given a policy $\mathcal{P}$, let us denote by $TGD(\mathcal{P})$ the following set of TGDs:

$$\{\forall \mathbf{x}_1, \mathbf{x}_2 \, (q_b(\mathbf{x}_1, \mathbf{x}_2) \rightarrow q_h(\mathbf{x}_2)) \mid \forall \mathbf{x}_1, \mathbf{x}_2 \, (Kq_b(\mathbf{x}_1, \mathbf{x}_2) \rightarrow Kq_h(\mathbf{x}_2)) \in \mathcal{P}\}.$$

Moreover, given a DL-Lite$_\mathcal{R}$ TBox $\mathcal{T}$, let us denote by $TGD(\mathcal{T})$ the set of TGDs obtained from $\mathcal{T}$ in the natural way. More details are given in the Appendix.

UCQ-rewritability is a well-established property of DL-Lite$_\mathcal{R}$ TBoxes [10], i.e.:

Proposition 2. *Let $\mathcal{T}$ be a DL-Lite$_\mathcal{R}$ TBox. Then, $TGD(\mathcal{T})$ is UCQ-rewritable.*

For the sake of readability, when referring to DL-Lite$_\mathcal{R}$ TBoxes $\mathcal{T}$, we write $\mathsf{UCQRew}(q, \mathcal{T})$ instead of $\mathsf{UCQRew}(q, TGD(\mathcal{T}))$.

Furthermore, let us indicate with $\mathcal{P}^+$ the policy containing every ED in $\mathcal{P}$ whose head does not contain $\bot$, and with $\Sigma(\mathcal{P}, \mathcal{T})$ the set of TGDs $TGD(\mathcal{P}^+) \cup TGD(\mathcal{T})$. We say that $\mathcal{P}$ is *expandable* w.r.t. $\mathcal{T}$ if $\Sigma(\mathcal{P}, \mathcal{T})$ is UCQ-rewritable. Then, from the results in [20], we directly have what follows:

Proposition 3. *Let $\mathcal{T}$ be a DL-Lite$_\mathcal{R}$ TBox, and let $\mathcal{P}$ be a policy that is either linear or acyclic for $\mathcal{T}$. Then, $\mathcal{P}$ is expandable w.r.t. $\mathcal{T}$.*[4]

We can now introduce Algorithm 1, for which we prove the following property.

Lemma 1. *Let $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$ be a DL-Lite$_\mathcal{R}$ CQE instance such that $\mathcal{P}$ is full and expandable w.r.t. $\mathcal{T}$, and let $\mathcal{F}$ be a set of facts. Then, $\mathcal{F}$ is disclosable in $\mathcal{E}$ iff:*

(i) $\mathcal{F} \subseteq \mathsf{cl}_\mathcal{T}(\mathcal{A})$ *and*
(ii) *for every* $\tau \in \mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})$ *and every ground substitution* σ *of the free variables of* $\mathsf{body}(\tau)$, $\mathcal{F} \models \sigma(\mathsf{body}(\tau))$ *implies that* $\mathcal{T} \cup \mathcal{A} \models \sigma(\mathsf{head}(\tau))$.

Proof. ($\Rightarrow$) Suppose that $\mathcal{F}$ is disclosable in $\mathcal{E}$. By Definition 4, this means that there exists a subset $\mathcal{F}'$ of $\mathsf{cl}_\mathcal{T}(\mathcal{A})$ such that $\mathcal{F} \subseteq \mathcal{F}'$ and $\mathcal{T} \cup \mathcal{F}' \models_{\mathsf{EQL}} \mathcal{P}$. Then, condition (i) directly follows from the fact that $\mathcal{F} \subseteq \mathcal{F}' \subseteq \mathsf{cl}_\mathcal{T}(\mathcal{A})$.

Let us now consider any τ and σ as in condition (ii), and let $\mathcal{F} \models \sigma(\mathsf{body}(\tau))$. To prove the *only-if* direction of the thesis, it remains to show that $\mathcal{T} \cup \mathcal{A} \models \sigma(\mathsf{head}(\tau))$. First, by construction of $\mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})$, there exists an ED $\tau' \in \mathcal{P}$

[3] Actually, that paper focuses on BCQs. Anyway, as also stated by the authors, the algorithm provided is easily extendable to open CQs.

[4] It is not hard to see that this property is preserved in the case the policy also contains denials. Thus, our technique applies to a generalization of the case of denials.

such that $\mathsf{body}(\tau) \in \mathsf{UCQRew}(\mathsf{body}(\tau'), \Sigma))$ (where Σ is as in Algorithm 1) and $\mathsf{head}(\tau') = \mathsf{head}(\tau)$. Consequently, we have that $\mathcal{F} \models \mathsf{UCQRew}(\mathsf{body}(\tau'), \Sigma))$ and, since Σ is UCQ-rewritable, that $\Sigma \cup \mathcal{F} \models \sigma(\mathsf{body}(\tau'))$.

Now, by construction of $\mathcal{F}'$ and since $\mathcal{P}^+$ is full, for every BCQ q, we have that $\mathcal{T} \cup TGD(\mathcal{P}^+) \cup \mathcal{F} \models q$ only if $\mathcal{T} \cup \mathcal{F}' \models q$. In particular, $\Sigma \cup \mathcal{F} \models \sigma(\mathsf{body}(\tau'))$ implies that $\mathcal{T} \cup \mathcal{F}' \models \sigma(\mathsf{body}(\tau'))$. Consequently, since $\mathcal{T} \cup \mathcal{F}' \models_{\mathsf{EQL}} \mathcal{P}$, we have that $\mathcal{T} \cup \mathcal{F}' \models \sigma(\mathsf{head}(\tau'))$. But since $\mathcal{F}' \subseteq \mathsf{cl}_{\mathcal{T}}(\mathcal{A})$, then by monotonicity it follows that $\mathcal{T} \cup \mathsf{cl}_{\mathcal{T}}(\mathcal{A}) \models \sigma(\mathsf{head}(\tau'))$, which holds iff $\mathcal{T} \cup \mathcal{A} \models \sigma(\mathsf{head}(\tau'))$. Finally, from the fact that $\mathsf{head}(\tau') = \mathsf{head}(\tau)$, we have that $\mathcal{T} \cup \mathcal{A} \models \sigma(\mathsf{head}(\tau))$.

($\Leftarrow$) Let now $\mathcal{F}$ be a set of facts satisfying conditions (i) and (ii).

Let us take the sequences of EDs $\tau_1, \ldots, \tau_m$, substitutions $\sigma_1, \ldots, \sigma_m$ and sets of facts $\mathcal{F}_1, \ldots, \mathcal{F}_{m+1}$ (with $\mathcal{F}_1 = \mathcal{F}$) such that, for every $1 \leq i \leq m$:

- $\tau_i \in \mathcal{P}$;
- σ_i is a ground substitution of the free variables of $\mathsf{body}(\tau_i)$ such that $\mathcal{F}_i \models \sigma_i(\mathsf{body}(\tau_i))$;
- $\mathcal{F}_{i+1} = \mathcal{F}_i \cup QA(\sigma_i(\mathsf{head}(\tau_i)))$.

Note that, since $\mathcal{P}$ is full, it is possible to choose m as a finite positive integer such that $\mathcal{F}_m$ consists of all the facts that are logical consequences of $TGD(\mathcal{P}) \cup \mathcal{F}$. Let us then set m to this upper bound.

We now show by induction that $\mathcal{F}_i \subseteq \mathsf{cl}_{\mathcal{T}}(\mathcal{A})$ for every $1 \leq i \leq m$. To this aim, let $\mathcal{P}_0 = \emptyset$ and, for every $1 \leq i \leq m$, let $\mathcal{P}_i$ be the set $\mathcal{P}_{i-1} \cup \{\tau_i\}$. Observe that, by construction, we have that $TGD(\mathcal{P}_{i-1}) \cup \mathcal{F} \models \sigma_i(\mathsf{body}(\tau_i))$ for every $1 \leq i \leq m$.

◊ The base case is trivial, as $\mathcal{F}_1 = \mathcal{F}$ is contained in $\mathsf{cl}_{\mathcal{T}}(\mathcal{A})$ by condition (i).

◊ For the inductive step, suppose that $\mathcal{F}_{i-1} \subseteq \mathsf{cl}_{\mathcal{T}}(\mathcal{A})$. We first show that $\tau_j \in \mathcal{P}^+$ for every $1 \leq j \leq i$. Towards a contradiction, let us consider the lowest j for which τ_j is a denial (i.e. $\mathcal{P}_{j-1} = \mathcal{P}^+_{j-1}$). In this case, we would have that $TGD(\mathcal{P}^+_{j-1}) \cup \mathcal{F} \models \sigma_j(\mathsf{body}(\tau_j))$, implying by monotonicity that $TGD(\mathcal{P}^+) \cup \mathcal{F} \models \sigma_j(\mathsf{body}(\tau_j))$, and hence that $\Sigma \cup \mathcal{F} \models \sigma_j(\mathsf{body}(\tau_j))$ (where Σ is as in Algorithm 1). Since Σ is UCQ-rewritable, then there would exist a CQ $q \in \mathsf{UCQRew}(\mathsf{body}(\tau_j), \Sigma)$ having the same free variables of $\mathsf{body}(\tau_j)$ such that $\mathcal{F} \models \sigma_i(q)$. Moreover, by Algorithm 1, $\mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})$ contains the ED $\tau = \forall \mathbf{x}\, (Kq(\mathbf{x}) \rightarrow K\mathsf{head}(\tau_i))$. Then, condition (ii) would imply that $\mathcal{T} \cup \mathcal{A} \models \sigma_j(\mathsf{head}(\tau_j))$, i.e. $\mathcal{T} \cup \mathcal{A} \models \bot$, which would contradict the hypothesis that the ontology $\mathcal{T} \cup \mathcal{A}$ is consistent.
From the fact that $\tau_j \in \mathcal{P}^+$ for every $1 \leq j \leq i$ it follows that $TGD(\mathcal{P}^+) \cup \mathcal{F} \models \sigma_i(\mathsf{body}(\tau_i))$. Thus, following the same reasoning as above, one can conclude $\mathcal{T} \cup \mathcal{A} \models \sigma_i(\mathsf{head}_i)$ (i.e. $\sigma_i(\mathsf{head}_i) \subseteq \mathsf{cl}_{\mathcal{T}}(\mathcal{A})$), which, together with the inductive hypothesis, implies that $\mathcal{F}_i \subseteq \mathsf{cl}_{\mathcal{T}}(\mathcal{A})$.

Finally, in order to prove that $\mathcal{T} \cup \mathcal{F}_m \models_{\mathsf{EQL}} \mathcal{P}$, let us take any ED $\tau \in \mathcal{P}$ and any ground substitution σ of the free variables of τ, and let us show that if $\mathcal{T} \cup \mathcal{F}_m \models \sigma(\mathsf{body}(\tau))$ then $\mathcal{T} \cup \mathcal{F}_m \models \sigma(\mathsf{head}(\tau))$. Suppose then that $\mathcal{T} \cup \mathcal{F}_m \models \sigma(\mathsf{body}(\tau))$, which by monotonicity holds only if $\Sigma \cup \mathcal{F}_m \models \sigma(\mathsf{body}(\tau))$. Since Σ is

UCQ-rewritable, there exists a CQ $q \in \mathsf{UCQRew}(\mathsf{body}(\tau), \Sigma)$, with the same free variables as $\mathsf{body}(\tau)$, such that $\mathcal{F}_m \models \sigma(q)$. Notice that $\mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})$ contains the ED $\tau' = \forall \mathbf{x}\, (Kq(\mathbf{x}) \to K\mathsf{head}(\tau))$. Then, by the construction of $\mathcal{F}_m$, from $\mathcal{F}_m \models \sigma(q)$ (that is, $\mathcal{F}_m \models \sigma(\mathsf{body}(\tau')))$ it follows that $QA(\sigma(\mathsf{head}(\tau'))) \subseteq \mathcal{F}_m$, which in turn implies that $\mathcal{T} \cup \mathcal{F}_m \models \sigma(\mathsf{head}(\tau'))$ and, since $\mathsf{head}(\tau') = \mathsf{head}(\tau)$, we conclude that $\mathcal{T} \cup \mathcal{F}_m \models \sigma(\mathsf{head}(\tau))$.

Thus, $\mathcal{F}_m$ is the set that meets the two conditions of Definition 4, that is, $\mathcal{F}$ is disclosable in $\mathcal{E}$. □

In the following, given a CQ q and a set of atoms $\mathcal{Z}$, we say that q is *mappable to $\mathcal{Z}$ via μ* if there exists a substitution μ replacing the variables of $\mathcal{Z}$ and the free variables of q with terms of $\mathcal{Z}$ and constants of q, in such a way that there exists a substitution μ' of the existential variables of q such that $\mu'(\mu(QA(q))) \subseteq \mu(\mathcal{Z})$. We indicate as $\mathsf{map}(q, \mathcal{Z})$ the set of all the substitutions μ such that q is mappable to $\mathcal{Z}$ via μ. E.g. if $\mathcal{Z} = \{R(x, y), R(z, w)\}$ and $q(v, u) = \exists t\, R(v, 1) \wedge R(u, t)$, then both $\{v \mapsto x, y \mapsto 1, u \mapsto z\}$ and $\{v \mapsto z, w \mapsto 1, u \mapsto x\}$ belong to $\mathsf{map}(q, \mathcal{Z})$.

Definition 5. *Given a DL-Lite$_\mathcal{R}$ TBox $\mathcal{T}$, a policy $\mathcal{P}$ expandable w.r.t. $\mathcal{T}$, and a set $\mathcal{Z}$ of atoms (we assume w.l.o.g. that every $x \in vars(\mathcal{Z})$ does not occur in* $\mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})$*), we define the formula* $\mathsf{isDiscl}(\mathcal{Z}, \mathcal{T}, \mathcal{P})$ *as follows:*

$$\mathsf{AtomRewr}(\mathsf{conj}(\mathcal{Z}), \mathcal{T}) \wedge \bigwedge_{\substack{\tau \in \mathsf{PolicyExp}(\mathcal{P}, \mathcal{T}) \,\wedge \\ \mu \in \mathsf{map}(\mathsf{body}(\tau), \mathcal{Z})}} \Big(\mathsf{eq}(\mu, \mathcal{Z}) \to \mathsf{UCQRew}(\mu(\mathsf{head}(\tau)), \mathcal{T})\Big)$$

where:

- $\mathsf{conj}(\mathcal{Z}) = \bigwedge_{\alpha \in \mathcal{Z}} \alpha$, *for every set $\mathcal{Z}$ of atoms;*
- $\mathsf{eq}(\mu, \mathcal{Z}) = \bigwedge_{\substack{x \mapsto t \in \mu \,\wedge \\ x \in vars(\mathcal{Z})}} x = t$ *(with true in place of the empty conjunction).*

Intuitively, for every ground substitution σ of its free variables, the FO sentence $\sigma(\mathsf{isDiscl}(\mathcal{Z}, \mathcal{T}, \mathcal{P}))$ encodes the two conditions of Lemma 1. This property is formally established by the next statement.

Lemma 2. *Let $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$ be a DL-Lite$_\mathcal{R}$ CQE instance such that $\mathcal{P}$ is full and expandable w.r.t. $\mathcal{T}$. Then, for every set of atoms $\mathcal{Z}$ and for every ground substitution σ of the variables of $\mathcal{Z}$, $\sigma(\mathsf{isDiscl}(\mathcal{Z}, \mathcal{T}, \mathcal{P}))$ evaluates to true in $\mathcal{A}$ iff $\sigma(\mathcal{Z})$ is disclosable in $\mathcal{E}$.*

Proof. ($\Leftarrow$) Let $\sigma(\mathcal{Z})$ be disclosable in $\mathcal{E}$. By Lemma 1, we have that:

(i) $\sigma(\mathcal{Z}) \subseteq \mathsf{cl}_\mathcal{T}(\mathcal{A})$ and
(ii) for every $\tau \in \mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})$ and every ground substitution σ' of the free variables of $\mathsf{body}(\tau)$, $\sigma(\mathcal{Z}) \models \sigma'(\mathsf{body}(\tau))$ implies that $\mathcal{T} \cup \mathcal{A} \models \sigma'(\mathsf{head}(\tau))$.

Clearly, from condition (i) it follows that $\sigma(\mathsf{conj}(\mathcal{Z}))$ evaluates to true in $\mathsf{cl}_\mathcal{T}(\mathcal{A})$. By Proposition 1, we then have that $\mathsf{AtomRewr}(\sigma(\mathsf{conj}(\mathcal{Z})), \mathcal{T})$ (which coincides with $\sigma(\mathsf{AtomRewr}(\mathsf{conj}(\mathcal{Z}), \mathcal{T})))$ evaluates to true in $\mathcal{A}$.

Now, let us consider any ED $\tau \in \mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})$ and any substitution $\mu \in \mathsf{map}(\mathsf{body}(\tau), \mathcal{Z})$. By definition, μ can be partitioned in two disjoint substitutions, namely μ_1 and μ_2, such that:

- μ_1 is a substitution of the free variables of $\mathsf{body}(\tau)$ with terms of $\mathcal{Z}$ and constants of $\mathsf{body}(\tau)$ such that $\mu(\mathsf{body}(\tau)) = \mu_1(\mathsf{body}(\tau))$;
- μ_2 is a substitution of the variables of $\mathcal{Z}$ with terms of $\mathcal{Z}$ and constants of $\mathsf{body}(\tau)$ such that $\mu(\mathcal{Z}) = \mu_2(\mathcal{Z})$;
- there exists a substitution μ' of the existential variables of $\mathsf{body}(\tau)$ (i.e. the ones of $\mu_1(\mathsf{body}(\tau))$) such that $\mu'(\mu_1(QA(\mathsf{body}(\tau)))) \subseteq \mu_2(\mathcal{Z})$.

Consider now the conjunction of atoms $\sigma(\mathsf{eq}(\mu, \mathcal{Z}))$, which, by definition of $\mathsf{eq}(\cdot, \cdot)$, is equal to $\sigma(\mathsf{eq}(\mu_2, \mathcal{Z}))$. As it is ground, it can either be unsatisfiable or valid (i.e. a conjunction of reflexive equalities on constants). In the first case, it obviously evaluates to false in $\mathcal{A}$. In the second case, instead, we have that by applying σ to any variable x of $\mathcal{Z}$ or to the term which x is mapped to via μ_2 one obtains the same constant, which implies that $\sigma(\mu_2(\mathcal{Z})) = \sigma(\mathcal{Z})$. Moreover, since σ does not replace existential variables of $\mu_1(QA(\mathsf{body}(\tau)))$, then by construction of μ' we have that $\mu'(\sigma(\mu_1(QA(\mathsf{body}(\tau))))) \subseteq \sigma(\mu_2(\mathcal{Z}))$, which holds only if $\sigma(\mathcal{Z}) \models \sigma(\mu_1(\mathsf{body}(\tau)))$. Now, note that the substitution σ' resulting by applying first μ_1 and then σ to $\mathsf{body}(\tau)$ is a ground substitution of the free variables of $\mathsf{body}(\tau)$. Therefore, by condition (ii), it follows that $\mathcal{T} \cup \mathcal{A} \models \sigma(\mu_1(\mathsf{head}(\tau)))$ (or equivalently $\mathcal{T} \cup \mathsf{cl}_{\mathcal{T}}(\mathcal{A}) \models \sigma(\mu_1(\mathsf{head}(\tau)))$). Then, by Proposition 2, the sentence $\sigma(\mathsf{UCQRew}(\mu_1(\mathsf{head}(\tau)), \mathcal{T}))$ (i.e. $\sigma(\mathsf{UCQRew}(\mu(\mathsf{head}(\tau)), \mathcal{T}))$) evaluates to true in $\mathcal{A}$.

Thus, for every $\tau \in \mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})$ and for every $\mu \in \mathsf{map}(\mathsf{body}(\tau), \mathcal{Z})$, either $\sigma(\mathsf{eq}(\mu, \mathcal{Z}))$ evaluates to false in $\mathcal{A}$ or $\sigma(\mathsf{UCQRew}(\mu(\mathsf{head}(\tau)), \mathcal{T}))$ evaluates to true in $\mathcal{A}$, from which the thesis immediately follows.

($\Rightarrow$) Now, suppose that $\sigma(\mathsf{isDiscl}(\mathcal{Z}, \mathcal{T}, \mathcal{P}))$ evaluates to true in $\mathcal{A}$. Then:

(i) $\sigma(\mathsf{AtomRewr}(\mathsf{conj}(\mathcal{Z}), \mathcal{T}))$ evaluates to true in $\mathcal{A}$, which by Proposition 1 implies that $\sigma(\mathsf{conj}(\mathcal{Z}))$ evaluates to true in $\mathsf{cl}_{\mathcal{T}}(\mathcal{A})$ (i.e. $\sigma(\mathcal{Z}) \subseteq \mathsf{cl}_{\mathcal{T}}(\mathcal{A})$);

(ii) for every $\tau \in \mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})$ and for every $\mu \in \mathsf{map}(\mathsf{body}(\tau), \mathcal{Z})$), if the (ground) conjunction $\sigma(\mathsf{eq}(\mu, \mathcal{Z}))$ evaluates to true in $\mathcal{A}$ then also the BUCQ $\sigma(\mathsf{UCQRew}(\mu(\mathsf{head}(\tau)), \mathcal{T}))$ does.

Let us consider any ED $\tau \in \mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})$ and any ground substitution σ' of the free variables of $\mathsf{body}(\tau)$ such that $\sigma(\mathcal{Z}) \models \sigma'(\mathsf{body}(\tau))$. For proving the thesis, we have to show that $\mathcal{T} \cup \mathcal{A} \models \sigma'(\mathsf{head}(\tau))$, which by Lemma 1 would imply that $\sigma(\mathcal{Z})$ is disclosable in $\mathcal{E}$.

First, since σ' replaces the free variables of $\mathsf{body}(\tau)$ with constants of $\sigma(\mathcal{Z})$, then there exists a substitution μ_1 of the free variables of $\mathsf{body}(\tau)$ with terms of $\sigma(\mathcal{Z})$ such that $\sigma'(\mathsf{body}(\tau)) = \sigma(\mu_1(\mathsf{body}(\tau))$. Therefore, it holds that $\sigma(\mathcal{Z}) \models \sigma(\mu_1(\mathsf{body}(\tau)))$, i.e. $\sigma(\mathcal{Z}) \supseteq \mu'(\sigma(\mu_1(QA(\mathsf{body}(\tau)))))$ for some substitution μ' of the existential variables of $\sigma(\mu_1(\mathsf{body}(\tau)))$ (i.e. of the ones of $\mathsf{body}(\tau)$). Then, one can see that there exists a substitution μ_2 of the variables of $\mathcal{Z}$ with terms of $\mathcal{Z}$ and constants of $\mathsf{body}(\tau)$ such that $\sigma(\mathcal{Z}) = \sigma(\mu_2(\mathcal{Z}))$ and $\mu_2(\mathcal{Z}) \supseteq$

$\mu'(\mu_1(QA(\mathsf{body}(\tau))))$. Intuitively, μ_2 is a weakened version of σ that allows not to apply σ on the right-hand side while preserving the homomorphic relationship between the two sets.

Now, let μ_1 and μ_2 be two substitutions such that what above holds, and observe that they replace distinct variables. Then, it is easy to see that the combination μ of μ_1 and μ_2 belongs to $\mathsf{map}(\mathsf{body}(\tau), \mathcal{Z})$. Moreover, since $\sigma(\mathcal{Z}) = \sigma(\mu_2(\mathcal{Z}))$, then we have that $\sigma(\mathsf{eq}(\mu_2, \mathcal{Z}))$ (which by construction is equivalent to $\sigma(\mathsf{eq}(\mu, \mathcal{Z})))$ is valid and, consequently, it evaluates to true in $\mathcal{A}$. Therefore, by condition (ii), we have that $\sigma(\mu(\mathsf{UCQRew}(\mathsf{head}(\tau), \mathcal{T})))$ evaluates to true in $\mathcal{A}$. By Proposition 2, this implies that $\mathcal{T} \cup \mathcal{A} \models \sigma(\mu(\mathsf{head}(\tau)))$. Then, the thesis follows by observing that $\sigma(\mu(\mathsf{head}(\tau))) = \sigma(\mu_1(\mathsf{head}(\tau))) = \sigma'(\mathsf{head}(\tau))$. □

Then, given a DL-Lite$_\mathcal{R}$ TBox $\mathcal{T}$, a policy $\mathcal{P}$ expandable w.r.t. $\mathcal{T}$, a set $\mathcal{Z}$ of atoms, and a CQ $q(\mathbf{x})$ without existential variables (we assume w.l.o.g. that $\mathbf{x} \cap \mathit{vars}(\mathcal{Z}) = \emptyset$), we define the FO formula $\mathsf{Clash}(\mathcal{Z}, q, \mathcal{T}, \mathcal{P})$ as follows:

$$\mathsf{Clash}(\mathcal{Z}, q, \mathcal{T}, \mathcal{P}) = \exists \mathbf{y}\ \big(\mathsf{isDiscl}(\mathcal{Z}, \mathcal{T}, \mathcal{P}) \land \neg \mathsf{isDiscl}(\mathcal{Z} \cup QA(q), \mathcal{T}, \mathcal{P})\big),$$

where $\mathbf{y}$ is a tuple containing all the variables occurring in $\mathcal{Z}$. Note that the variables of $\mathbf{x}$ are free. In words, by existentially closing the formula $\mathsf{Clash}(\mathcal{Z}, q, \mathcal{T}, \mathcal{P})$ and then evaluating it over an ABox $\mathcal{A}$ it is possible to check whether there exists a common instantiation σ for all the atoms of $\mathcal{Z}$ and q such that (i) the set $\sigma(\mathcal{Z})$ is disclosable (and thus it is part of some censor) and (ii) it is no longer disclosable when we add the atoms of q. Intuitively, if there exists an instantiation of the variables of q for which the above does not occur for any $\mathcal{Z}$, then q is entailed by $\mathcal{E}$ under the IGA semantics.

In order to obtain a proper FO-rewriting algorithm, it remains to show that the size of $\mathcal{Z}$ can be upper-bounded by a certain integer k that is independent of $\mathcal{A}$. To this aim, given a set of predicates $\Pi = \{p_1, \ldots, p_m\}$ and a positive integer k, we define the following set of atoms:

$$\mathit{Atoms}(\Pi, k) = \{p(\mathbf{x}_i) \mid p \in \Pi \text{ and } i \in \{1, \ldots, k\}\},$$

where each $\mathbf{x}_i$ is a sequence of h fresh variables, if h is the arity of p.

We are now ready to provide the following FO-rewriting function.

Definition 6. *Let $\mathcal{T}$ be a DL-Lite$_\mathcal{R}$ TBox, let $\mathcal{P}$ be a policy expandable w.r.t. $\mathcal{T}$, and let q be a BUCQ. Then* $\mathsf{IGA\text{-}Ent}(q, \mathcal{T}, \mathcal{P})$ *is the sentence:*

$$\mathsf{IGA\text{-}Ent}(q, \mathcal{T}, \mathcal{P}) = \bigvee_{\exists \mathbf{x}\, \gamma(\mathbf{x}) \in q_r} \exists \mathbf{x} \Big(\gamma \land \bigwedge_{\substack{\mathcal{Z} \subseteq \mathit{Atoms}(\mathit{pred}(\mathcal{P} \cup \mathcal{T}), k) \\ \land |\mathcal{Z}| < k}} \neg \mathsf{Clash}(\mathcal{Z}, \gamma, \mathcal{T}, \mathcal{P})\Big).$$

where $q_r = \mathsf{UCQRew}(q, \mathcal{T})$ *and* $k = \max_{\tau \in \mathsf{PolicyExp}(\mathcal{P}, \mathcal{T})} |QA(\mathsf{body}(\tau))|$.

It is possible to prove that, for every DL-Lite$_\mathcal{R}$ CQE specification $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$, the sentence $\mathsf{IGA\text{-}Ent}(q, \mathcal{T}, \mathcal{P})$ evaluates to true in $\mathcal{A}$ iff $\mathcal{E} \models_{\mathsf{IGA}} q$. Such a FO-rewritability property implies the next theorem.

Theorem 4. *Let $\mathcal{E} = \langle \mathcal{T}, \mathcal{P}, \mathcal{A} \rangle$ be DL-Lite$_\mathcal{R}$ CQE instance such that $\mathcal{P}$ is full and expandable w.r.t. $\mathcal{T}$, and let q be a BUCQ. Deciding whether $\mathcal{E}$ IGA-entails q is in* AC^0 *w.r.t. data complexity.*

Proof. Let $\mathcal{A}$ be any ABox, let $q_i = \exists \mathbf{x}\, \gamma(\mathbf{x})$ be the BCQ such that γ occurs in the i-th disjunct of $\mathsf{IGA\text{-}Ent}(q, \mathcal{T}, \mathcal{P})$, let $\mathcal{Z}$ be a set of atoms, and let σ be an instantiation of q_i in $\mathcal{A}$. Then, it is immediate to verify that $\sigma(\mathsf{Clash}(\mathcal{Z}, \gamma, \mathcal{T}, \mathcal{P}))$ evaluates to true in $\mathcal{A}$ iff there exists an instantiation σ' of $\mathsf{conj}(\mathcal{Z})$ in $\mathcal{A}$ such that $\mathsf{isDiscl}(\sigma'(\mathcal{Z}), \mathcal{T}, \mathcal{P})$ and $\mathsf{isDiscl}(\sigma'(\mathcal{Z}) \cup \sigma(QA(q_i)), \mathcal{T}, \mathcal{P})$ evaluate, respectively, to true and false in $\mathcal{A}$. Consequently, by Lemma 2, we have the following property:

> **(PR1)**: The sentence $\sigma(\mathsf{Clash}(\mathcal{Z}, q_i, \mathcal{T}, \mathcal{P}))$ evaluates to true in $\mathcal{A}$ iff there exists an instantiation σ' of $\mathsf{conj}(\mathcal{Z})$ in $\mathcal{A}$ such that $\sigma'(\mathcal{Z})$ is disclosable in $\mathcal{E}$ and $\sigma'(\mathcal{Z}) \cup \sigma(QA(q_i))$ is not.

The next statement follows from the previous one and from the fact that, for every set of facts $\mathcal{F} \subseteq \mathcal{A}$ of size not greater than a given integer k', there exists a subset $\mathcal{Z}$ of $Atoms(pred(\mathcal{P} \cup \mathcal{T}), k')$ such that $\mathcal{F}$ is an image of $\mathsf{conj}(\mathcal{Z})$ in $\mathcal{A}$.

> **(PR2)**: Let ϕ be the i-th disjunct of $\mathsf{IGA\text{-}Ent}(q, \mathcal{P}, \mathcal{T})$. Then ϕ evaluates to true in $\mathcal{A}$ iff there exists an image M of q_i in $\mathcal{A}$ such that, for every set of facts $\mathcal{F}$ such that $\mathcal{F} \subseteq \mathcal{A}$ and $|\mathcal{F}| < k$ (where k is as in Definition 6), if $\mathcal{F}$ is disclosable in $\mathcal{E}$ then also $\mathcal{F} \cup M$ is.

Furthermore, the fact that $\mathcal{P}$ is expandable w.r.t. $\mathcal{T}$ implies that, if there exists a set $\mathcal{F}$ of facts such that $\mathcal{F}$ is disclosable in $\mathcal{E}$ and $\mathcal{F} \cup M$ is not, then there exists a set of facts $\mathcal{F}'$ such that $\mathcal{F}'$ is disclosable in $\mathcal{E}$, $\mathcal{F}' \cup M$ is not and $|\mathcal{F}'| < k$.

This last property, along with (PR2), implies that q_i evaluates to true in $\mathcal{C}_{\mathsf{IGA}}$ iff the i-th disjunct of $\mathsf{IGA\text{-}Ent}(q, \mathcal{T}, \mathcal{P})$ evaluates to true in $\mathcal{A}$. Note in fact that, for every instantiation σ of q_i in $\mathcal{A}$, the above set $\mathcal{F}'$ of facts exists iff $\sigma(QA(q_i))$ is not contained in at least one optimal GA censor for $\mathcal{E}$ (i.e. the one containing $\mathcal{F}'$). Therefore, by Proposition 2 and since $q_i \in \mathsf{UCQRew}(q, \mathcal{T})$, we have that $\mathcal{T} \cup \mathcal{C}_{\mathsf{IGA}} \models q$ (i.e. $\mathcal{E} \models_{\mathsf{IGA}} q$) iff $\mathsf{IGA\text{-}Ent}(q, \mathcal{T}, \mathcal{P})$ evaluates to true in $\mathcal{A}$, which proves that BUCQ entailment under IGA semantics is FO-rewritable, and thus in AC^0 w.r.t. data complexity. □

Example 4. Recalling Example 2, let us verify that $\mathcal{E} \not\models_{\mathsf{IGA}} q_1$ by rewriting it and then evaluating it over the ABox. First, observe that $\mathsf{UCQRew}(q_1, \mathcal{T}) = q_1$, i.e. the TBox does not affect the entailment of q_1 in this case, and that the unique instantiation of q_1 in $\mathcal{A}$ is $\sigma = \{x \mapsto \mathsf{lucy}, y \mapsto \mathsf{tom}\}$. In addition, since $k = 2$, all sets $\mathcal{Z}$ of Definition 6 are singletons. In particular, for the set $\mathcal{Z} = \{\mathsf{managerOf}(x', y')\}$ one can see that, under the assignment $\sigma' = \sigma \cup \{x' \mapsto \mathsf{lucy}, y' \mapsto \mathsf{tom}\}$, the sentence $\sigma'(\mathsf{Clash}(\mathcal{Z}, q_1, \mathcal{T}, \mathcal{P}))$ evaluates to true in $\mathcal{A}$. Therefore, the whole sentence $\mathsf{IGA\text{-}Ent}(q_1, \mathcal{T}, \mathcal{P})$ evaluates to false in $\mathcal{A}$.

For q_2, we have $\mathsf{UCQRew}(q_2, \mathcal{T}) = \{q_2, \exists x, z(\mathsf{managerOf}(x, z) \wedge \mathsf{salary}(x, \mathsf{150k})), \exists x\, (\mathsf{manager}(x) \wedge \mathsf{salary}(x, \mathsf{150k}))\}$. In particular, one can verify that for the last BCQ there exists no set $\mathcal{Z}'$ analogous to the above set $\mathcal{Z}$ for query q_1, which implies that q_2 is IGA-entailed by $\mathcal{E}$. ■

7 Experiments

In this section, we describe the experiments that we conducted to test the feasibility of our approach. We evaluated our queries on a standard laptop with an Intel i7-8565U @1.8 GHz processor and 16GB of RAM.

We refer to the OWL2Bench benchmark for OWL ontologies [24], which models the university domain and includes a tool for generating ABoxes of customizable size (measured as number of universities). For our experiments, we tested all the ten SPARQL queries for OWL 2 QL against the $\mathsf{o2b}_5$ and $\mathsf{o2b}_{10}$ ABoxes, which store data about 5 and 10 universities (i.e. ∼325k and ∼710k ABox assertions), respectively. We removed the data properties from the input ontology (as they are not part of DL-Lite$_\mathcal{R}$), and added the definition of the `knows` object property and the `Woman` class in the TBox for OWL 2 QL[5], as both such predicates occur in the generated ABoxes. The ABoxes were stored in an SQL database, as the FO query produced by our rewriting algorithm can be naturally translated into SQL. We were thus able to delegate the evaluation of queries on the ABox to the SQL system.

As for the reasoner, we employed the tree-witness query rewriter for OWL2 QL ontologies[6] [19,23]. It can be used as an actual implementation of the UCQRew abstract rewriting function when its second argument is a DL-Lite$_\mathcal{R}$ TBox.

The two cases on which we focused our experiments are the ones in which the policy is either full and linear or full and acyclic for the coupled TBox. More precisely, for the first case we define a slightly more restricted language of EDs (defined below), which we call *binary EDs*. The correspondence of such dependencies with DL-Lite$_\mathcal{R}$ axioms allows us to exploit the tree-witness rewriter in place of UCQRew for rewriting the EDs' bodies. For the case of acyclic policies, instead, we implemented a specific version of UCQRew from scratch.

Although Theorem 4 constitutes a remarkable theoretical result, from a practical point of view, the size of the rewritten formula $\mathsf{IGA\text{-}Ent}$ may have a severe impact on the evaluation time. For this reason, we made several intensional optimizations in our implementation to obtain a simpler yet semantically equivalent rewriting. The most important are the following ones.

- Instead of generating a Clash subformula for every possible $\mathcal{Z}$ set defined in $\mathsf{IGA\text{-}Ent}$, we can only consider those sets that, for a fixed $q \in q_r$, can match at least one body of the expanded policy.
- In the $\mathsf{isDiscl}$ subformula, the atoms of the query can be omitted from $\mathsf{conj}(\mathcal{Z})$ (at this point of the evaluation, we know that their rewriting is satisfied). Then, since the two $\mathsf{AtomRewr}(\mathcal{Z}, \mathcal{T})$ in Clash become syntactically equal, they can only be evaluated once, that is, outside the two $\mathsf{isDiscl}$.
- For its specific purpose, the output of the map function can be optimized in terms of specificity of the substitutions. Specifically, we do not use a substitution μ such that q is mappable to $\mathcal{Z}$ via μ if we also use a substitution μ' that is more generic than μ and such that q is mappable to $\mathcal{Z}$ via μ'.

[5] https://github.com/kracr/owl2bench/blob/master/UNIV-BENCH-OWL2QL.owl
[6] https://titan.dcs.bbk.ac.uk/~roman/tw-rewriting/

Our system's source code can be downloaded from the url: https://github.com/anonymous-iswc25/iga-ed-rewriter/blob/main/iga-ed-rewriter.zip.

7.1 Binary EDs

We now define the fragment of binary EDs, a subclass of linear EDs which has a correspondence with DL-Lite$_\mathcal{R}$.

Definition 7. *A* binary ED *is an ED of one of the following forms:*

$$\forall x \, (KB_1 \rightarrow KB_2) \tag{2}$$
$$\forall x, y \, (KS_1 \rightarrow KS_2) \tag{3}$$

where B_1 and B_2 are expressions of one of the following forms:

$$\{A(x),\ \exists y \, R(x, y),\ \exists y \, R(y, x)\}$$

(where A is a concept name and R_1, R_2 are role names), and S_1 and S_2 are expressions of one of the following forms:

$$\{R(x, y),\ R(y, x)\}$$

We now define the function $DL(\cdot)$ as follows:

$$DL(A(x)) = A \qquad DL(\exists y \, R(y, x)) = \exists R^- \qquad DL(R(x, y)) = R$$
$$DL(\exists y \, R(x, y)) = \exists R \qquad DL(R(y, x)) = R^-$$

Then, given a binary ED τ of the form (2), we define $DL(\tau)$ as the DL-Lite$_\mathcal{R}$ concept inclusion $DL(B_1) \sqsubseteq DL(B_2)$, and given a binary ED τ of the form (3), we define $DL(\tau)$ as the DL-Lite$_\mathcal{R}$ role inclusion $DL(S_1) \sqsubseteq DL(S_2)$. Finally, if $\mathcal{P}$ is a set of binary EDs, we define $DL(\mathcal{P})$ as the DL-Lite$_\mathcal{R}$ TBox $\bigcup_{\tau \in \mathcal{P}} DL(\tau)$.

As said above, this kind of policy enabled us not to re-implement a rewriting function UCQRew of Algorithm 1 for linear EDs. Formally, when $\mathcal{P}$ is a binary policy, instead of $\mathsf{UCQRew}(\mathsf{body}(\tau), \Sigma(\mathcal{P}, \mathcal{T}))$, we computed $\mathsf{UCQRew}(\mathsf{body}(\tau), DL(\mathcal{P}^+) \cup \mathcal{T})$ by exploiting the tree-witness rewriter also for this purpose.

One may wonder whether this restriction is sufficient to get back FO-rewritability. However, the kind of policy used in the proof of Theorem 2 immediately rules out this possibility. Hence, we restrict our attention to the case of *full* binary EDs, which is FO-rewritable by Theorem 4.[7]

[7] We recall that the restriction to a single atom in the head does not actually decrease the expressiveness of full EDs (and thus of full binary EDs).

7.2 Policy Definition

We ran our experiments using an acyclic policy $\mathcal{P}_{\mathrm{a}}$ (consisting of 6 EDs) and a binary policy $\mathcal{P}_{\mathrm{b}}$ (consisting of 11 EDs), namely (we removed the universally quantified variables for improving their readability):

$$
\begin{array}{l}
\{\, K\,\exists x\,\mathsf{isAdvisedBy}(x,y) \rightarrow K\,\mathsf{Woman}(y),\\
\;\; K\,\exists x\,(\mathsf{hasAlumnus}(x,y) \land \mathsf{hasMasterDegreeFrom}(y,x) \rightarrow K\,\mathsf{FullProfessor}(y),\\
\;\; K\,(\mathsf{takesCourse}(x,y) \land \mathsf{Student}(x)) \rightarrow K\,\bot,\\
\;\; K\,(\mathsf{hasCollaborationWith}(x,y) \land \mathsf{Student}(x)) \rightarrow K\,\bot,\\
\;\; K\,\mathsf{isAdvisedBy}(x,y) \land \mathsf{hasMasterDegreeFrom}(x,z) \rightarrow K\,\mathsf{hasMajor}(x,\mathsf{ComputerScience}),\\
\;\; K\,(\mathsf{teachesCourse}(x,y) \land \mathsf{FullProfessor}(x)) \rightarrow K\,\mathsf{hasDoctoralDegreeFrom}(x,\mathsf{U2})\,\},\\
\{\, K\,\mathsf{isAffiliatedOrganizationOf}(x,y), \mathsf{hasCollegeDiscipline}(x,\mathsf{FineArts}) \rightarrow K\,\bot,\\
\;\; K\,\mathsf{Professor}(x) \rightarrow K\,\mathsf{Woman}(x),\\
\;\; K\,\mathsf{teachesCourse}(x,y) \rightarrow K\,\mathsf{FullProfessor}(x),\\
\;\; K\,\mathsf{isVisitingProfessorOf}(x,y) \rightarrow K\,\bot,\\
\;\; K\,\mathsf{takesCourse}(x,y) \rightarrow K\,\mathsf{ElectiveCourse}(y),\\
\;\; K\,\mathsf{Person}(x) \rightarrow K\,\mathsf{Employee}(x),\\
\;\; K\,\mathsf{hasAlumnus}(x,y) \rightarrow K\,\mathsf{hasMasterDegreeFrom}(y,x),\\
\;\; K\,\mathsf{hasCollaborationWith}(x,y) \rightarrow K\,\mathsf{Professor}(x),\\
\;\; K\,\mathsf{hasSameHomeTownWith}(x,y) \rightarrow K\,\mathsf{Employee}(x),\\
\;\; K\,\mathsf{knows}(x,y) \rightarrow K\,\mathsf{Professor}(x),\\
\;\; K\,\mathsf{knows}(x,y) \rightarrow K\,\mathsf{Professor}(y)\,\}.
\end{array}
$$

We also defined $\mathcal{P}_{\mathrm{a}}^{-}$ as a reduced version of $\mathcal{P}_{\mathrm{a}}$, containing its last 3 EDs, and $\mathcal{P}_{\mathrm{b}}^{-}$ as a reduced version of $\mathcal{P}_{\mathrm{b}}$, containing its last 6 EDs.

7.3 Results

The results of our experiments are reported in Table 1. The symbol $\mathcal{P}_{\emptyset}$ indicates the absence of a policy: this is the configuration that we used as a baseline.

The main indications provided by these results are the following:

- In most cases, the evaluation time t_e is acceptable (it remains on the order of seconds). The only critical query is the seventh one, which takes several minutes to execute.
- The time t_r necessary for computing the rewritten query is always less than or equal to 3 seconds. As expected, for both $\mathsf{o2b}_5$ and $\mathsf{o2b}_{10}$, such rewriting times are very close, as they do not depend on the ABox.
- For both acyclic and binary policies, the values of t_r corresponding to smaller and larger policies are of comparable magnitude.
- Observe that binary policies tend to remove more tuples than acyclic ones. This could be explained by the fact that EDs with fewer atoms in their body are more likely "activated" by the query.

Query		o2b$_5$					o2b$_{10}$				
		$\mathcal{P}_\emptyset$	$\mathcal{P}_{\mathrm{a}}^{-}$	$\mathcal{P}_{\mathrm{a}}$	$\mathcal{P}_{\mathrm{b}}^{-}$	$\mathcal{P}_{\mathrm{b}}$	$\mathcal{P}_\emptyset$	$\mathcal{P}_{\mathrm{a}}^{-}$	$\mathcal{P}_{\mathrm{a}}$	$\mathcal{P}_{\mathrm{b}}^{-}$	$\mathcal{P}_{\mathrm{b}}$
q_1	t_r	460	60	58	169	193	690	64	87	303	212
	t_e	513	526	547	683	958	822	560	964	1394	1085
	#	9228	9228	9228	1367	334	19782	19782	19782	2948	730
q_2	t_r	50	249	336	476	511	65	243	386	873	553
	t_e	81	8688	7335	174	206	269	39426	51181	876	402
	#	18872	18736	14829	5957	5957	44190	43889	33193	13009	13009
q_3	t_r	25	41	42	40	109	38	32	43	62	112
	t_e	4	8	6	2	4	8	5	9	5	6
	#	34	34	34	34	28	75	75	75	75	64
q_4	t_r	21	40	35	52	133	33	33	49	68	97
	t_e	6	7	2	4	3	7	5	7	5	3
	#	0	0	0	0	0	0	0	0	0	0
q_5	t_r	21	554	473	250	319	29	482	917	451	334
	t_e	17	30699	29679	1823	1003	44	110763	124305	5056	1283
	#	3574	2020	2020	952	264	6564	3676	3676	1696	394
q_6	t_r	18	114	115	161	116	25	148	109	263	193
	t_e	59	60661	28564	81	230	235	283020	141173	333	282
	#	16236	15834	7811	0	0	35889	35075	17481	0	0
q_7	t_r	75	2029	1976	1378	1847	88	2205	2244	2222	1801
	t_e	66	198641	187908	403	496	334	993701	1119538	1343	812
	#	5489	5489	5091	5489	3292	11969	11969	10971	11969	7241
q_8	t_r	22	52	48	257	263	21	49	44	399	272
	t_e	62	75	63	615	647	186	112	143	1548	963
	#	17904	17904	17904	14668	14668	39278	39278	39278	32350	32350
q_9	t_r	135	1711	2302	1765	3430	195	1778	2290	2877	3144
	t_e	31	1412	1313	137	129	93	10230	10944	598	332
	#	1698	1539	1539	0	0	3434	3196	3196	0	0
q_{10}	t_r	22	1243	1220	184	337	32	1351	1433	380	477
	t_e	78	676	3	110	159	297	4753	3	478	270
	#	642	122	0	642	144	1413	258	0	1413	335

Table 1. All the evaluation results. Each entry reports the rewriting time (t_r) and evaluation time (t_e) expressed in milliseconds, plus the number of returned tuples (#).

8 Conclusions

In this work, we investigated CQE under policies expressed via epistemic dependencies (EDs), focusing on the use of ground atom (GA) censors for safe information disclosure. We investigated IGA-entailment, a semantic relation based on the intersection of all optimal GA censors, and analyzed its data complexity in the presence of ED-based policies when the TBox is expressed in DL-Lite$_\mathcal{R}$. Our results show that the intersection remains safe for full EDs, a subclass of particular interest. We established that IGA-entailment is not FO-rewritable in general, which led us to the definition of the subclass of full and expandable EDs for which FO-rewriting is feasible, and we introduced a rewriting algorithm that works for DL-Lite$_{\mathcal{R}}$ ontologies. We validated our approach through a prototype implementation evaluated using the OWL2Bench benchmark, showing its practical feasibility in diverse evaluation scenarios.

As for future work, we are currently working on the definition of practical algorithms for some of the non-FO-rewritable cases of IGA-entailment analyzed in the paper. Another interesting research direction is towards extending the FO-rewritable cases identified by our analysis: in particular, we would like to focus either on further subclasses of EDs, or on (subclasses of) policy languages that go beyond the expressiveness of EDs. In addition, it would be interesting to study the complexity of CQE under EDs for TBoxes expressed in DLs different from DL-Lite$_{\mathcal{R}}$. Finally, it would be of practical importance to extend the present approach to the framework of ontology-based data access (OBDA).

References

1. Abiteboul, S., Hull, R., Vianu, V.: Foundations of Databases. Addison Wesley Publ. Co. (1995)
2. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, 2nd edn. (2007)
3. Baget, J., Leclère, M., Mugnier, M., Salvat, E.: On rules with existential variables: Walking the decidability line. Artificial Intelligence **175**(9-10), 1620–1654 (2011)
4. Biskup, J.: For unknown secrecies refusal is better than lying. Data and Knowledge Engineering **33**(1), 1–23 (2000)
5. Biskup, J., Bonatti, P.A.: Controlled query evaluation for enforcing confidentiality in complete information systems. Int. J. Inf. Sec. **3**(1), 14–27 (2004)
6. Bonatti, P.A.: A false sense of security. Artificial Intelligence **310**(103741) (2022)
7. Bonatti, P.A., Sauro, L.: A confidentiality model for ontologies. In: Proc. of the 12th Int. Semantic Web Conf. (ISWC). Lecture Notes in Computer Science, vol. 8218, pp. 17–32. Springer (2013)
8. Calì, A., Gottlob, G., Lukasiewicz, T.: A general Datalog-based framework for tractable query answering over ontologies. J. of Web Semantics **14**, 57–83 (2012)
9. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: EQL-Lite: Effective first-order query processing in description logics. In: Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI). pp. 274–279. Morgan Kaufmann Publishers Inc. (2007)
10. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. J. of Automated Reasoning **39**(3), 385–429 (2007)
11. Cima, G., Lembo, D., Marconi, L., Rosati, R., Savo, D.F.: Controlled query evaluation in ontology-based data access. In: The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part I. Lecture Notes in Computer Science, vol. 12506, pp. 128–146. Springer (2020)
12. Cima, G., Lembo, D., Marconi, L., Rosati, R., Savo, D.F.: Enhancing controlled query evaluation through epistemic policies. In: Proc. of the 33th Int. Joint Conf. on Artificial Intelligence (IJCAI). pp. 3307–3314. Int. Joint Conf. on Artificial Intelligence Organization (Aug 2024)

13. Cima, G., Lembo, D., Marconi, L., Rosati, R., Savo, D.F.: A gentle introduction to controlled query evaluation in DL-Lite ontologies. Springer Nature Computer Science **5**(4), 335 (2024)
14. Cima, G., Lembo, D., Marconi, L., Rosati, R., Savo, D.F.: Indistinguishability in controlled query evaluation over prioritized description logic ontologies. J. of Web Semantics **84**, 100841 (2025)
15. Cima, G., Lembo, D., Rosati, R., Savo, D.F.: Controlled query evaluation in description logics through consistent query answering. Artificial Intelligence **334**, 104176 (2024)
16. Console, M., Lenzerini, M.: Epistemic integrity constraints for ontology-based data management. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence (AAAI). pp. 2790–2797. AAAI Press (2020)
17. Cuenca Grau, B., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U.: OWL 2: The next step for OWL. J. of Web Semantics **6**(4), 309–322 (2008)
18. Cuenca Grau, B., Kharlamov, E., Kostylev, E.V., Zheleznyakov, D.: Controlled query evaluation over OWL 2 RL ontologies. In: Proc. of the 12th Int. Semantic Web Conf. (ISWC). pp. 49–65 (2013)
19. Kikot, S., Kontchakov, R., Zakharyaschev, M.: Conjunctive query answering with OWL 2 QL. In: Proc. of the 13th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR). AAAI Press (2012)
20. König, M., Leclère, M., Mugnier, M., Thomazo, M.: Sound, complete and minimal UCQ-rewriting for existential rules. Semantic Web J. **6**(5), 451–475 (2015)
21. Motik, B., Cuenca Grau, B., Horrocks, I., Wu, Z., Fokoue, A., Lutz, C.: OWL 2 Web Ontology Language profiles (second edition). W3C Recommendation, World Wide Web Consortium (Dec 2012), available at http://www.w3.org/TR/owl2-profiles/
22. Papadimitriou, C.H.: Computational Complexity. Addison Wesley Publ. Co. (1994)
23. Rodriguez-Muro, M., Kontchakov, R., Zakharyaschev, M.: Ontology-based data access: Ontop of databases. In: Proc. of the 12th Int. Semantic Web Conf. (ISWC). Lecture Notes in Computer Science, vol. 8218, pp. 558–573. Springer (2013)
24. Singh, G., Bhatia, S., Mutharaju, R.: OWL2Bench: A benchmark for OWL 2 reasoners. In: Proc. of the 19th Int. Semantic Web Conf. (ISWC). Lecture Notes in Computer Science, vol. 12507, pp. 81–96. Springer (2020)